Pergamon

0042-6989(95)00298-7

Vision Res., Vol. 36, No. 16, pp. 2553–2562, 1996

Printed in Great Britain
0042-6989/96 $15.00 + 0.00

Saturation Revealed by Clamping the Gain of the Retinal Light Response

CHRISTOPHER W. TYLER,*† LEI LIU*

Received 29 January 1993; in revised form 13 March 1995; in final form 3 October 1995

The saturation nonlinearity of the retinal light response in human was measured by a psychophysical technique in which the adaptive gain control mechanism was clamped by the presence of a fixed surround in a small (7′) foveal test field. Gain clamping was established by showing that the normal variation in temporal summation properties with test intensity was abolished in the gain clamping paradigm. The static saturation function constructed from the increment/decrement asymmetries around a range of base intensities was shown to conform more closely to the Naka–Rushton hyperbolic saturation equation than to three other candidate nonlinearities. Copyright © 1996 Elsevier Science Ltd.

Saturation Luminance Distortion Gain control Temporal integration

INTRODUCTION

Nonlinearities can be valuable as guideposts in the investigation of any intact system, such as the human visual system or the physiological preparation of an isolated receptor cell. Once a nonlinearity is established, it can be used to determine which processes occurred before it and which after it. If the physiological site of the nonlinearity is known, the processes can then be partitioned relative to that site.

One well-established nonlinearity of neural responses is the saturation of the output signal of the receptors at high response levels. This nonlinearity takes the form of a gradual saturation to an asymptotic level, approximating the form of either the hyperbolic equation of Michaelis–Menton first applied to retina by Naka & Rushton (1966), or the asymptotic exponential equation of Baylor *et al.* (1987). Our aim was to reveal this static saturation nonlinearity and to determine its analytic form in a psychophysical paradigm for human vision. To do so, we needed to deactivate the adaptation mechanism that normally adjusts the response output level so as to avoid the output saturation region.

It is well known that when the adaptation field is large and static the cone system never saturates and the contrast threshold increases as the background intensity increases (Weber's law) up to the highest measurable light levels. Presumably, gain control is responsible for this proportionality up to the light level where pigment bleaching begins (Stiles, 1961; Shapley & Enroth-Cugell, 1986; Hood & Finkelstein, 1986). At this point, bleaching takes over, and threshold continues to conform to Weber's Law due to the filtering effect of the reduction of active pigment by the bleaching process. Note that the threshold increases due to gain control are, in principle, independent of those predicted from a saturating intensity function. In fact, if a fully effective gain control preceded the saturation process, it would hold the mean response level at a constant position on the saturation function as intensity was varied, making it impossible to measure the saturation characteristic.

Several attempts have been made to defeat such gain control. Hood *et al.* (1978) superimposed a small, brief probe upon a larger, longer duration flash. The flash used up the available dynamic range of the gain control mechanism and drove the system to nonlinear saturation behavior. This method confirmed that cones, like rods, could exhibit saturation. It also explored the temporal aspects of the saturation process. However, because of the transient nature of this saturation method, the spatial configuration of the stimulus was not important in producing saturation.

Taking an alternative approach, Buss *et al.* (1982) found that when the adaptation field was small enough, the cone threshold vs intensity curve exhibited a slope that was much steeper than that predicted by Weber's law. They interpreted this result to imply that the mechanism that controls the gain responsible for Weber's law integrates from an extended spatial area, so that a

*Smith–Kettlewell Eye Research Institute, 2232 Webster Street, San Francisco CA 94115, U.S.A.

†To whom all correspondence should be addressed [*Fax* +1 415 561 1610; *Email* cwt@skivs.ski.org *Internet* http//www.ski.org/cwt.html

small pedestal sitting at the center of large surround would have a very weak effect on the gain of the cone system. As a result, the gain would be determined largely by the intensity of the surround and allow the saturation nonlinearity to be probed by setting its level by the small adaptation field.

If it is true that the adaptive gain control has a large spatial spread then we should be able to show that (i) if the surround intensity changed the gain would change; and (ii) when the surround was constant the gain would always take the same value and be independent of the pedestal level.

Gain, defined by Shapley & Enroth-Cugell (1986) as "the ratio of the magnitude of the physiological response to the stimulus magnitude", can be measured directly through physiological recording, as has been done on the cone systems of a variety of different animal species. However, psychophysical measurement of gain has not been as successful, partially because the psychophysical threshold is determined not only by the gain but also by the inherent noise. However, some physiological studies on invertebrate ommatidia (Fuortes & Hodgkin, 1964) and the vertebrate cone system (Baylor & Hodgkin, 1974) revealed a close relation between the gain and the time course of cones, allowing us to infer the operation of gain control under appropriate conditions if the same relation holds for human vision. Baylor & Hodgkin (1974) found that, as the steady background increased, the gain of red cones in turtle retina decreased and time course of the response also speeded up (i.e., the time-to-peak shortened and the width of the impulse response became narrower). Daly & Normann (1985) showed that these changes in response speed of turtle cones mirror the temporal response of human flicker sensitivity functions with intensity (de Lange, 1958; Kelly, 1961), implying that different gain levels have different temporal integration properties, which therefore can be used to evaluate whether gain control is operating. If the gain for a small test area is held constant by a surround field of fixed intensity, then the temporal response should remain invariant as the base intensity level for the test stimulus is varied. Conversely, if the gain is varied by varying the surround level, the response speed for a small test area should change, even though the test stimulus intensity is held constant. [The fact that similar gain changes are not reported in direct current recordings from monkey (Baylor *et al.*, 1987) or human (Schnapf *et al.*, 1987) cones, which raises difficulties in interpreting the site of the gain mechanism, will be addressed in the Discussion.]

The aim of the study was to measure the saturation response of a single cone pathway in human vision. The long wavelength sensitive (R-cone) pathway was selectively stimulated by use of deep red light in a small foveolar spot (cf. Tyler & Hamer, 1987). With the surround at a fixed intensity, thresholds for both increments and decrements were measured from a variety of pedestal intensities of the foveolar spot. The threshold values then were used to construct the saturation function in effect under these conditions, on the assumption that no dynamic gain control was in operation. This assumption was supported by tests of the response dynamics.

GENERAL METHODS

Apparatus

The flash stimulus was presented by an array of 25 Stanley Ultra-Superbrite light-emitting diodes (LEDs) with a peak wavelength of 660 nm and a bandwidth at half height of 20 nm. (The dominant wavelength was 642 nm.) One of the advantages of these LEDs was that they could easily deliver increments and decrements of different amplitudes from a base level, which was required in the design of this experiment. The LED array was set behind a diffusing sheet in a tube with a white inner surface, so as to give the appearance of a uniform field. A tiny aperture was put in front of the LEDs to make a circular test field 7′ in diameter. The field was set in the center of a large (40 cm) hemispheric white surround, uniformly illuminated by a steady incandescent light source. The intensity of the surround could be set to different levels through an Acvar rheostat (Fig. 1).

The LED intensity was controlled by an Apple II + computer through a high-resolution D/A converter. For each threshold, a steady intensity pedestal (base level, I_B) was set at some level above or below that of the white surround. From the base, intensity threshold was measured for an increment or decrement probe, I_P, consisting of the whole area of the test field. The test field also served as a pretest adaptation field with an intensity equal to the base level when there was no stimulus.

The D/A output/luminance function of the LED array was measured with a UDT photometer. The light output for the 660 nm LEDs was not a linear function of the input current, but rose with calibrated power of 1.111 over the 2.4 log unit range of the D/A. The nonlinearity of this function was compensated for by incorporating the calibration function into the output algorithm, resulting in a linear control of light output. The time course of the

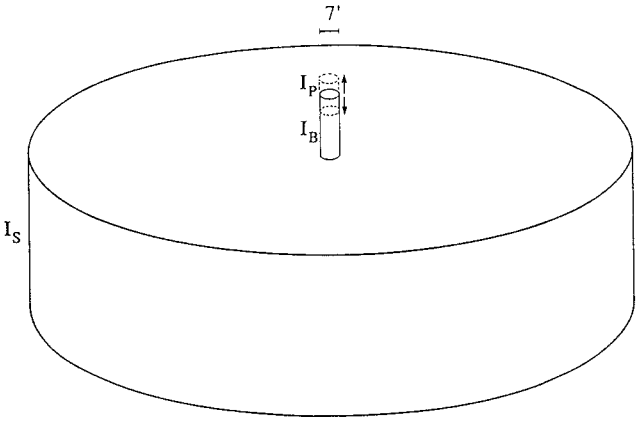

FIGURE 1. Stimulus dimensions. The stimulus consisted of a 660 nm test pulse, I_P , from a base pedestal, I_B , of the same diameter (7′), embedded in a full-field equiluminant white surround, I_B.

stimulus, duration, number of pulses in each stimulus and interpulse interval were also controlled by the computer to the precision of 1 msec. The intensity of the surround was calibrated with a Spectraspot light meter. By the use of neutral density filters, the range of measurements was extended to about 4 log units. These neutral density filters were calibrated by inserting them in the view path of the Spectraspot meter when in the position corresponding to the observer's eye and viewing the target at the test wavelength. They could be put either in front of the LEDs to lower the base level but not the surround, or in goggles worn by the observer to reduce both the surround and the base level.

For Experiment I, the observer was dilated (to a 9 mm pupil), both to achieve a high illuminance range and to eliminate pupil compensation for the variations in surround illuminance. Potentially, such dilation could have blurred the stimulus to such an extent that the contrast of the pedestal was substantially attenuated by the spread of the surround into the pedestal region. To evaluate the degree of blurring, we measured the modulation threshold function under the test conditions with and without dilation. At the 10% contrast level, spatial resolution was reduced from 16 to 13 c/deg but there was no detectable attenuation of contrast sensitivity at low spatial frequencies. Visibility of the 7′ stimulus, therefore, should have been little affected since the frequency most relevant for seeing the 7′ stimulus is only 4 c/deg (the test grating whose bars are each 7.5 deg in width), although the border would have appeared somewhat more blurred than without dilation. We therefore are confident that the pedestal level effectively controlled the retinal illuminance in the 7′ test field for the dilated as well as the undilated condition.

Psychophysical procedure

The observers fixated the stimulus monocularly from a chin-rest at 25 cm from the stimulus aperture. Before starting the measurement, the eye to be tested was exposed to the surround intensity for at least 1 min in order to become adapted. (Because most surround intensities used were higher than or almost equal to the ambient level, 1 min was enough for the eye to reach a steady light adaptation level.) The psychophysical thresholds were collected using a constant stimulus method in which increments and decrements were randomly intermixed. The observer first ran a staircase procedure to determine the proper mean stimulus level for the current testing condition and then started a testing procedure with five stimulus levels centered at the mean stimulus, separated by 0.1 log unit steps.

We used a parameter searching algorithm to find the best-fitting Weibull function:

$$W(c) = 1 - (1 - \gamma)e^{-\left(\frac{c}{\alpha}\right)^{\beta}} \qquad (1)$$

for the frequency-of-seeing curve (Weibull, 1951). The linear estimation of the slope of the function was used as the initial estimate of β_0. The initial threshold α_0 was calculated from β_0 and the overall false alarm rate γ using the equation:

$$\alpha_0 = [1 - (1 - \gamma)e - A]/\beta_0 \qquad (2)$$

The derivation of the equation is in the Appendix. Centered at the initial α_0 and β_0, a 5×5 array of α and β values was generated, in steps of 0.1 log unit. Weibull functions for each combination of α and β in the array were calculated and compared with the data. If the cell that held the least-squares error between the estimated Weibull function and the measured data was one of the cells on the periphery of the array, it became the center of a new array. The procedure was repeated until the cell with the least-square error was one of the inner nine cells of the array, implying that it was surrounded by cells of higher values, when the search process was terminated and the α value of the cell used as the threshold. This procedure was an effective means of searching the two-parameter space, on the assumption that the least-squares fit had only one minimum value.* Typically between one and three iterations were required to obtain the final value from the initial linear estimate of β.

EXPERIMENT I: CLAMPING THE GAIN

Ultimately, we wanted to measure the saturating nonlinearity of the cone pathway, which determines the response properties of the signal when the threshold increment or decrement responses are large relative to the base response level. To reveal this nonlinearity is difficult because the retinal adaptation processes set the gain of the system such that the response remains in the linear range, to protect the system from a "saturation catastrophe". So, the aim of the first experiment was to provide evidence that a suitable spatial configuration can prevent the gain control from functioning.

To determine whether gain control was in operation, we compared the temporal integration properties of the cone system under conditions of normal gain control and conditions designed to clamp the local gain at a fixed level. To allow normal gain control [Fig. 2(A)], the surround in each condition was set equal to each level of base intensity of the test spot, while for the gain clamping conditions [Fig. 2(B)], the surround was held constant while the base level changed. Under each combination of base and surround, temporal integration was measured by presenting two successive increments, each with the same height and a duration of 5 msec, and determining thresholds for different interpulse intervals. Inferring the quantitative features of the impulse response function under the testing condition from the two-pulse summation function has been a standard approach and has been

* It is theoretically possible for the minimum to escape the net of this search procedure due to undersampling of the parameter space. We used sampling intervals of 0.1 log units for both α and β. We verified that this was an appropriate sampling density by analyzing sets of typical data with a variety of starting values. The algorithm always ended at the same value, regardless of the choice of initial values.

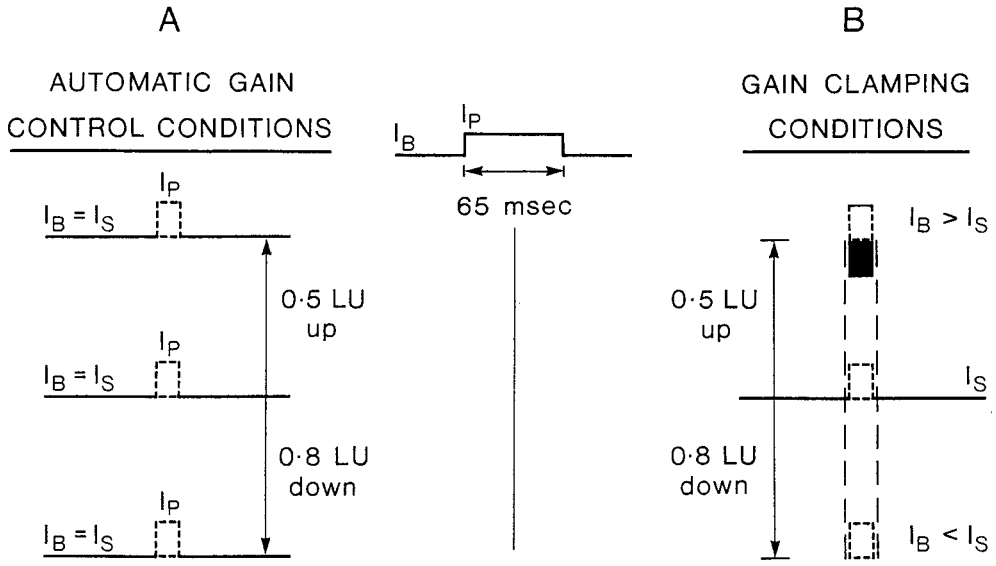

FIGURE 2. Stimulus intensity relations. Under normal conditions allowing the operation of gain control mechanisms (A), the white surround was set at the same three levels as the red pedestal, I_B. (The values shown for the pedestal levels are representative of those in the actual experiment.) The red test pulse was then incremented (or decremented) for 65 msec duration to determine thresholds. To achieve gain clamping (B), the same paradigm was used, except that the white surround was held at a fixed mean level (I_S) throughout so that the pedestal could be at either higher or lower intensity.

used by many researchers (Ikeda, 1965; Rashbass, 1970). Because it is the surround of a small field that is presumed to determine the adaptation level, the two-pulse response at any given surround intensity should be predictable from the two-pulse response at the same base intensity for a uniform background field. Although a systematic assay of two-pulse summation behavior with intensity was not conducted, good information is available about the temporal integration of pulses of varying duration from, for example, Graham & Kemp (1938). Their data show that the integration time varies by about a factor of two in the photopic luminance range, so that this is the order of variation we should expect for the two-pulse summation time over this intensity range.

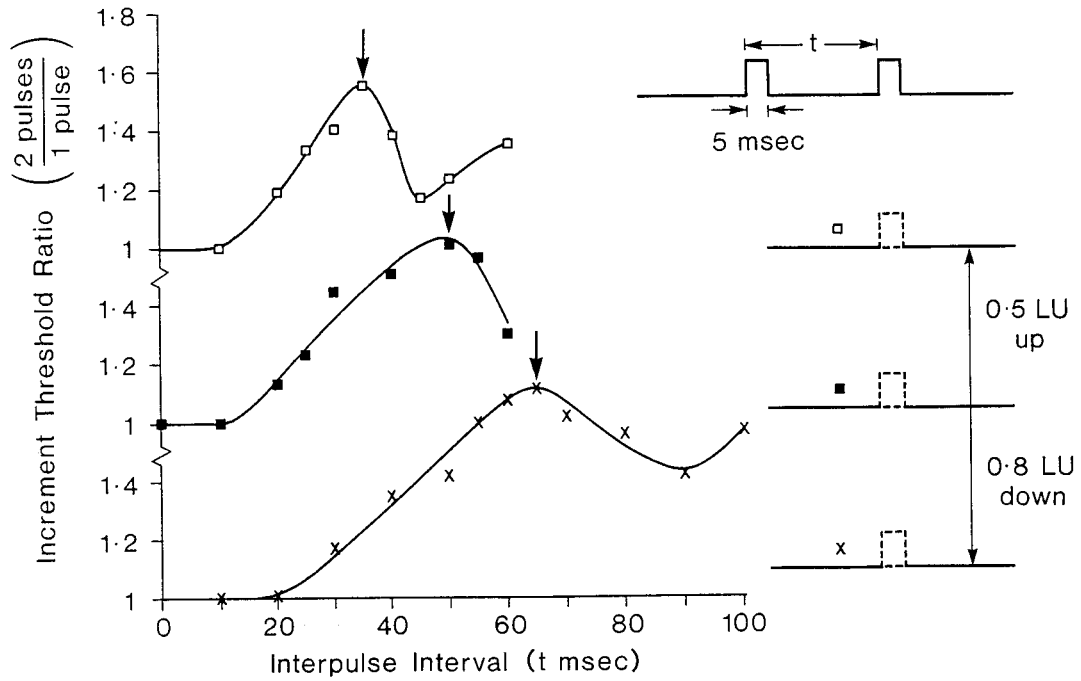

FIGURE 3. Two-pulse interactions under gain control operation, as specified in the bottom-right inset. The ordinate shows the ratio of increment thresholds for two 65 msec pulses with the specified interpulse interval (top-right inset) to the increment threshold for zero interpulse interval (single pulse). When the stimulus conditions allowed gain control to operate, the peak two-pulse summation occurred at shorter intervals (from 65 to 35 msec; arrows) as intensity increased. The mean standard error on each data set averaged 0.02 log units (about ± 0.05 around the measured ordinate ratios).

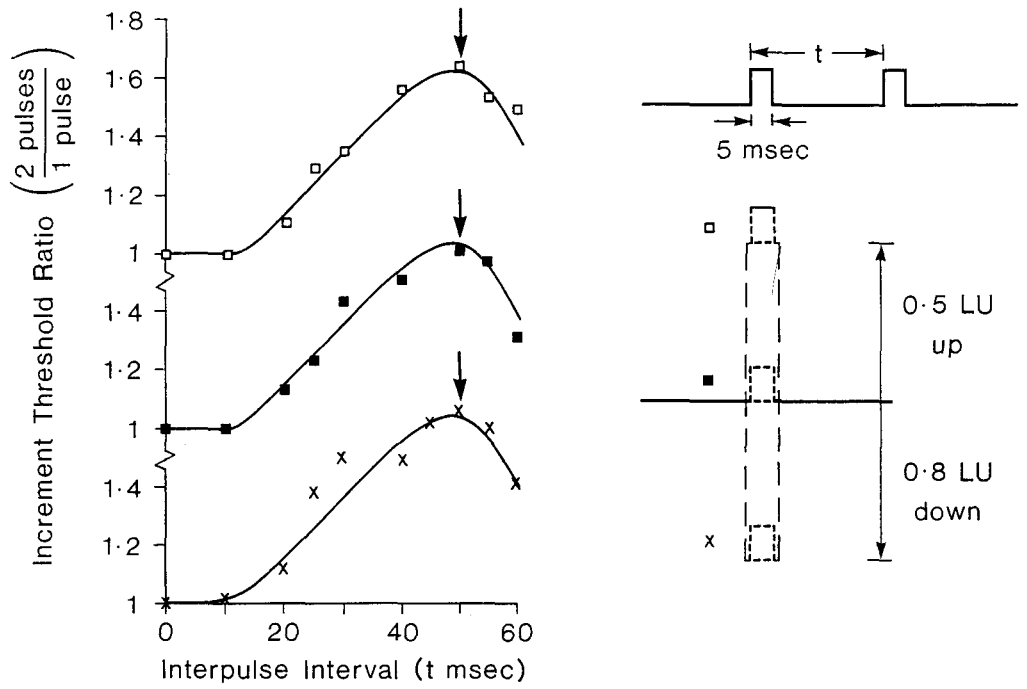

FIGURE 4. Gain-clamped two-pulse interaction. Similar to Fig. 3 except that the white surround was fixed to produce a gain clamping condition. The data at all three pedestal intensities are well fit by the same function, implying no change in the temporal integration properties with pedestal intensity (arrows). The mean standard error on each data set averaged 0.02 log units.

Operating gain control conditions

In this set of experiments, the intensity of the base level was set at three levels of 2523, 15,895 and 50,266 Td for observer LL, whose right eye was dilated throughout the experiment. At each illuminance level he first adjusted the surround so that it matched the base level in brightness. Increment thresholds for single and double increments then were determined with interpulse intervals from 1 to 100 msec. With very short interpulse intervals (less than 10 msec), the thresholds were comparable to twice the threshold for a single pulse.

Figure 3 shows the two-pulse summation functions for the three base (and surround) levels, with the data normalized to the single pulse threshold in each case. The ordinate shows the ratio of increment thresholds for two 5 msec pulses with the specified interpulse interval to the increment threshold for zero interpulse interval (twice the threshold for a single pulse). Deriving the shape of the impulse response from these curves is not the theme of this paper; a simple comparison of the curves under different surround intensities is sufficient for our purpose of establishing that the surround can influence the gain. As the surround level was increased, the peak of the curves (which indicates the point of failure of temporal integration) occurred earlier, and the time span over which the energy from the two pulses could be integrated (the width of the increment phase) shrank, together implying that the underlying impulse response function became faster and narrower. Since the shape of the response function covaried with the intensity level, we conclude that the change in time constant can be used as an indicator of the level of gain under our stimulus conditions.

Gain clamping conditions

Two-pulse summation thresholds were also measured under conditions where the surround was always set at 6150 Td while the base level was set equal to, and higher and lower than, the surround by the same amounts as in the operating gain conditions (namely, 2523, 15,895 and 50,266 Td). The data, normalized to single pulse thresholds, are plotted in Fig. 4. The fitted curves are drawn from the same template and show that the three sets of data all peaked at essentially the same interpulse interval, and that all other features of the psychophysical temporal interaction were similarly invariant to within the error of measurement. This invariance implies that all the temporal features of the underlying retinal response controlling the observer's behavior were similarly invariant. Thus, we see that, once the surround was fixed the temporal summation became invariant, supporting the idea that the different base levels had very little effect on the retinal gain, as though the gain of the small test area were effectively clamped by the surround.

EXPERIMENT II: MEASURING THE SATURATION NONLINEARITY

Under conditions where gain control operates, the gain is adjusted according to the adaptation intensity so that increment stimuli of the same contrast would always produce the same input on the same linear portion of the non-linearity, whatever the base adaptation level. That is

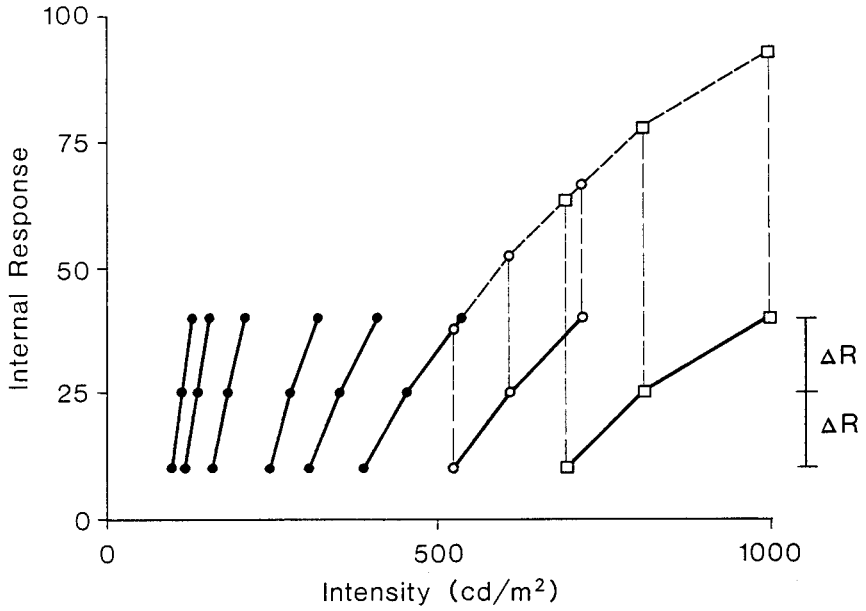

FIGURE 5. Construction of a saturation function from measured increment and decrement thresholds. Internal response of $\pm\Delta R$ is assumed to have an amplitude of 15 arbitrary units for increments and decrements around all base intensities, giving rise to a set of two-segment functions (solid lines) placed arbitrarily at an initial base response level of 25. The fitting procedure consists of vertically shifting the two-segment functions to derive the best-fitting overall function (thick dashes). The vertical shift was applied to each base response level according to the parameters of a candidate saturation function (see text).

why we can enjoy contrast constancy in daily life. If the gain control ceased to function, then pulses from different base levels would hit different parts of the output nonlinearity. In particular, when the base was high enough, the increment stimulus would start near the ceiling of response range and saturation would occur. Now that we are equipped with a technique to clamp the gain, we can use the pulses from different base levels to explore the whole dynamic range of the output function and reveal the saturation nonlinearity.

One of the most distinctive features of saturation nonlinearity is an asymmetry of increment and decrement responses. If there is an underlying saturation nonlinearity, a decrement from a base level should be more detectable than an increment from the same base. This asymmetry will become more pronounced when the base approaches the ceiling of the dynamic range. This property can be easily analyzed if we make some assumptions about the way the visual system responds to the stimuli.

Assumption 1: The internal responses to the base and to the pulse from the base are linearly additive, that is, the overall response is the sum of the response to the base and the response to the test probe ($R_B + R_P$).

Assumption 2: The internal response threshold, ΔR, is the same for incremental and decremental stimuli and this threshold response is constant at all base response levels.

Assumption 2 is not necessarily equivalent to assuming that threshold is limited by a fixed, additive internal noise. The rationale for a fixed (or high) threshold assumption, independent of internal noise is discussed and empirically supported for flashed intensity stimuli in Tyler (1991) [see also Klein & Tyler (1993)]. Modeling by Graham & Hood (1992) also shows that the deterministic, or fixed threshold model accounts for temporal psychophysical data better than models limited by quantal noise. The close equivalence of incremental and decremental responses for flashed stimuli is well established (see review in Tyler *et al.*, 1992) despite recent demonstrations of marked asymmetries for ramped (Bowen *et al.*, 1989) and smooth temporally modulated (Tyler *et al.*, 1992) stimuli.

By Assumption 1, the base intensity I_B sets the operating point of the response at R_B. If the internal responses required to reach threshold are both equal to ΔR (Assumption 2), then a larger increment is needed to induce this response due to the saturation. At this stage, the explicit expression of the saturation is indeterminate. It is difficult to find any psychophysical evidence to justify Assumption 1, but any distortions of the internal response metric from the assumed linearity will not invalidate the method, but only modify the form of the estimated saturation function according to the distortion of the metric.

In fact, Assumption 1, which Williams & Gale (1977) called "the additive model", is seldom true in general. As long as the gain control mechanism is able and has enough time to do its job, the gain for the test probe I_P will become adjusted according to I_B, so that the overall response $R_B + R_P$ will always fall short of the response elicited by a single flash of magnitude $I_B + I_P$. However,if one can inactivate the gain control, either by using a test probe following very closely to the onset of adaptation field I_P before the gain can be adjusted, as did Hood *et al.* (1978), or by using a tiny field sitting in the center of a large surround, as we propose, then this additivity may be valid. We therefore designed conditions to generate the required inactivation of the gain control nonlinearity, and

describe a test that validates the efficacy of these conditions. Given that gain control is inactivated, Assumption 1 should apply and allow reconstruction of the saturation function to proceed.

When we measured the increment and decrement thresholds, the experimental paradigm was to intermingle two constant stimulus processes, for increment and decrements from the same base level, and determine both thresholds in a single run. At threshold, in any single run, it was difficult to distinguish between an increment and a decrement. Although the pulse heights needed for eliciting a threshold response were different for increments and decrements, it seemed that the two types of perturbation, from some level on, shared the same processing mechanism.

Figure 5 illustrates how the up/down asymmetry caused by the saturation, can be used for reconstruction of the whole saturation behavior through the psychophysical experiment. Suppose that we obtain a pair of thresholds for an increment and a decrement from a certain base level. According to Assumptions 1 and 2, the response to the increment is ΔR up from the response to the base, whatever level the base is, and the decrement response is ΔR down. Using the threshold pulse heights and this knowledge, we can find a pair of points on the internal response axis. Connecting them and the point representing the base, we get a three-point function which has to be a part of the saturation curve. From each base level, we determine a pair of thresholds and draw a segment of the curve. All we have to do is to somehow connect them to make a smooth curve. Because the gain is clamped, these curve segments are locked by their base stimulus levels and cannot be shifted horizontally. The only free parameter is the response to the base, which is unknown. If we can shift the individual curve segments up and down so that they all lie on a single curve, this curve should represent the saturation nonlinearity we are investigating. The analysis of Experiment II follows this logic.

Psychophysics

To set two different gain levels, we measured the saturation nonlinearity for two surround levels of 6150 Td (undilated) for both observers and a dilated condition amounting to 50,266 Td for observer LL and 24,600 Td for observer AS (the difference resulting from their different pupil sizes when dilated). Under each surround level, we chose approximately ten base levels, covering about 2 log units. The actual values were 4.27, 3.33, 1.33, 1.00, 0.67, 0.5, 0.33, 0.23 and 0.17 times the surround level. These base levels were chosen so that the adjacent pieces of curve should have some overlap. This requirement, however, turned out to be unnecessary because every segment was almost self-contained. Once its position (the base level) was fixed, whether there was overlap or not had little effect on the later reconstruction process. For each base level, the thresholds for an increment and a decrement from this level were determined with the intermixed constant stimulus method. The observer first adapted his eye to the surround and then used an interleaved staircase procedure to determine the mean pulse levels for the up and down stimuli separately. The stimuli were presented in a sequence of increments and decrements with varying heights.

For threshold data of LL at the lower surround level (Fig. 5), the increment and decrement thresholds are displayed around the corresponding base level. The horizontal axis represents the base level and also shows the pulse heights needed to elicit a threshold response. The threshold response, ΔR, was arbitrarily chosen as 15 units, with the only restriction that ΔR for the increment and the decrement were equal.

Reconstructing the saturation nonlinearity

Figure 5 also shows schematically how we used the curve segments obtained from the psychophysical measurements to reconstruct the global curve of the response saturation. If Assumption 2 is true, the response to a pulse from a base is always the sum of the response to the pulse and the response to the base. Translating the curve segments vertically is equivalent to assigning some value to the base response, which is itself unmeasurable. The fitting procedure consists of varying this vertical shift in some way to minimize the deviation of the measured points from the resulting global curve. The approach we will take is to optimize several theoretical functions for the global curve to determine which provides the best-fit to the increment and decrement data. It will be seen that the shape of a segment and the intensity of the base (i.e., its horizontal location) impose a strong restriction on the shape of the global curve.

To assess quantitatively the function best describing the data, we started with a set of possible models:

– hyperbolic nonlinearity (Naka–Rushton) equation. The response, R, is given by:

$$R = a \cdot I/(I + b) \tag{3}$$

where I is the stimulus intensity, a is the gain and b is the half-saturation. These symbols denote corresponding constants in the subsequent equations:

– logarithmic compression,

$$R = a \cdot log(1 + bI) \tag{4}$$

– exponential saturation

$$R = a(1 - e^{bI}) \tag{5}$$

– power law,

$$R = a \cdot I^b \tag{6}$$

Each of these models has two free parameters. For each model, we used the same two-dimensional parameter searching algorithm as described for the psychometric functions to look for the parameter set best-fitting the data. A 5 × 5 array of cells was selected around the designated starting values and a curve calculated for the parameters in each cell. The base points on the curve segments from the data were then pinned at the computed curve and the threshold increment/decrement heights

TABLE 1. Comparison of saturation fits for four models

	Model	Optimal parameters a	b	Mean square error (% Range)
LL				
	$R = aI/(I + b)$	432	1538	0.31
	$R = a(1\text{-}e^{-I/b})$	330	5748	0.87
	$R = a\log(1 + I/b)$	244	627	0.49
	$R = aI^b$	1429	3.24	1.01
AS				
	$R = aI/(I + b)$	188	2050	1.18
	$R = a(1\text{-}e^{-bI})$	175	1.10	29.93
	$R = a\log(1 + bI)$	67.2	35732	31.71
	$R = aI^b$	119432	0.015	31.20

Best fitting parameter pairs for the four candidate saturation equations for each observer. Errors are expressed in terms of mean c^2 as a percent of the range of the data on the response axis for each fitted response.

TABLE 2. Hyperbolic saturation fits at two intensity levels

	Td	a	b	% Error
LL				
	6150	432	1538	0.21
	50266	354	9410	0.46
AS				
	6150	188	2050	0.97
	24600	271	30500	1.80

Best fitting parameter pairs for the hyperbolic saturation equation [equation (3)] for the two adaptation levels (in Td) for each observer. As distinct from Table 1, errors are expressed in terms of mean c^2 as a percent of the asymptotic saturation level, a, rather than the range of the data on the response axis. This change was introduced because the range of data collected was limited by the apparatus at high intensities. The relation may be appreciated from the values for the hyperbolic fit at 6150 Td, which are based on the identical fit in the two tables.

compared with the pulse heights derived from the curve. The squared errors in stimulus intensity were then averaged to serve as the estimate of goodness of fit for this particular cell. A coarse-to-fine scaling was employed to speed up the search while maintaining good precision. This algorithm worked well and converged to the same point over a large range of initial parameters.

The first application of this procedure was selecting the model which fitted best to our data. Table 1 compares the fit of the lower intensity (6150 Td) data with the four models for two observers. The hyperbolic equations gave the smallest least-square error for both observers, by an order of magnitude for AS. There may be better models which can yield even smaller errors, but of the four functions tested, the hyperbolic saturation equation provides the best description of our data.

We can compare the characteristics of our reconstructed saturation nonlinearity with those found in physiological and other psychophysical studies. First consider the exponent k in the Michaelis–Menten equation:

$$R = aI^k/(I^k + b^k) \quad (7)$$

In the literature of cone system, this parameter has been given values ranging from 0.7 to 1, but most researchers have agreed on a value of $k = 1$ (e.g., Baylor & Hodgkin, 1973 and Normann & Perlmann, 1979, in turtle; Boynton & Whitten, 1970 and Schnapf *et al.*, 1990, in monkey). In the study of cone adaptation, the same investigators have also found that k did not change with the adaptation level (gain). We therefore used a value of 1 for the two-parameter curve fits in Table 1, to hold the number of parameters the same for all four equations to be compared.

However, the optimal value of k can also be determined empirically by optimizing the hyperbolic saturation equation of equation (7) to the data. For this model, we used a three-dimensional parameter searching algorithm to look for the best-fitting parameter set. We took a cube from the three-dimensional parameter space and divided it to $5 \times 5 \times 5$ cells. A curve was calculated with the parameters from each cell and the optimal parameter set determined as in the two-dimensional search routine. The optimal values for k were found to be 1.04 and 1.05 for the two intensity levels for observer LL, and 0.86 and 1.17 for observer AS. The mean value of k for the two conditions was 1.045 for observer LL and 1.015 for observer AS, supporting a value of 1 as a good approximation.

For comparison between the two observers, two-parameter fits of the rectangular hyperbolic equation (with k set to 1) were now computed for all data sets. The parameter values and mean standard deviation (as a percentage of the asymptotic maximum value, a, in each case) are given in Table 2. For LL the root mean square errors were all less than 0.5% of the maxima, while for AS they were all less than 2%. These values represent excellent fits to the hyperbolic saturation equation at all levels. Moreover, the parameter values obtained for a were relatively similar, given that the function at the higher level is determined only over a limited range.

Many physiological and psychophysical studies on gain control have also revealed that the main effect of a changing gain is to change only the half-saturation parameter (e.g., Baylor & Hodgkin, 1974; Normann & Perlmann, 1979; Boynton & Whitten, 1970; Hood *et al.*, 1978). If plotted on a log–log scale, curves of decreasing gain levels will show a rightward horizontal shift. Figure 6 shows our reconstructed saturation nonlinearity plotted in the same way, showing a clear rightward shift of the curve as the surround intensity increased. There were also small changes in the saturation asymptote, a, at the highest intensity level by -0.12 and $+0.21$ log units for the two observers. However, Fig. 6 makes it evident that the parameter values for the highest level are derived by extended extrapolation from much lower levels, so the changes should not be regarded as significant.

DISCUSSION

Our results support the presence of a static nonlinearity in light processing under small-field conditions. The nonlinearity conforms better to the hyperbolic Naka–Rushton equation than to three alternative forms of the nonlinearity that have been suggested in the literature. Having established its form, the main question raised by

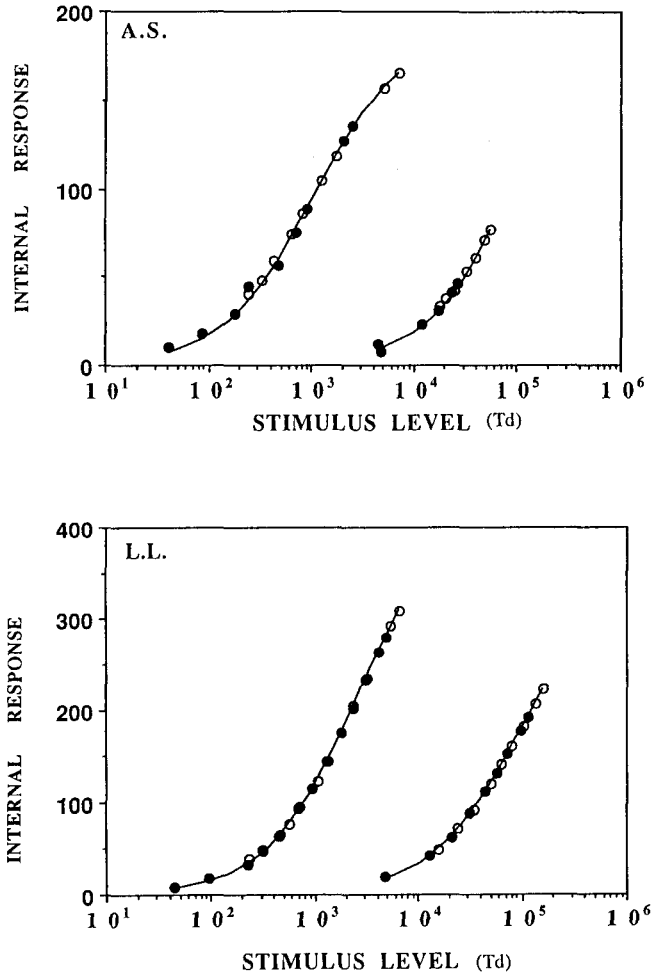

FIGURE 6. Log saturation functions under gain clamping conditions. Estimated two-parameter saturation functions on log–linear coordinates for the range of base levels at each surround level used, for two observers (see text). Open circles, increment thresholds; solid circles, decrement thresholds. Note similarity in shapes of saturation functions for each observer as they slide horizontally in response to the gain clamping effect of the surround.

these results is the site of the nonlinearity. There are two salient possibilities for this site—at the output of the photoreceptors and at a subsequent neural site, such as the output of the ganglion cells. Each possibility carries implications which are incompatible with data from the current literature in retinal signal processing; we have therefore been unable to exclude either one at present.

Before considering the two hypotheses for the site of the static nonlinearity, it is important to evaluate one possible organization suggested by recent recordings from monkey and human cones (Baylor *et al.*, 1987; Schnapf *et al.*, 1987, 1990). This suggestion is that there is little gain control preceding the nonlinearity we have measured. Neither monkey nor human cones in the isolated outer segment preparation show the extent of either gain control or time constant changes that would be expected from human psychophysics or from turtle cone recordings. If this were how the cones functioned *in vivo* then they would always be subject to the saturating nonlinearity and a gain clamping paradigm would not be required to reveal it. Varying the base intensity of the stimulus would then vary the level of the response on the output nonlinearity and incremental thresholds would no longer conform to Weber's law. In fact, as Williams & Gale (1977) showed by taking the derivative of the saturation function, thresholds would grow with the square of base intensity if the saturation function were hyperbolic [equation (1)]. The actual derivative of equation (3) is $ab/(I + b)^2$, but this expression needs to be inverted to model the threshold elevation function as measured psychophysically, giving the expression $(I + b)^2/ab$. Differentiating also shows that thresholds would grow with the power function $I^b - 1/ab$ if equation (4) described the saturation, and would grow exponentially according to e^{bI}/ab if the saturation function were exponential [equation (5)]. Only if the functions measured in Fig. 6 were logarithmic [equation (6)] would the threshold function conform to Weber's law [specifically, $(1 + bI)/ab$ for the form of equation (7)].

Since a large body of literature (e.g., Stiles, 1961) shows that the threshold/intensity function conforms to Weber's law under most conditions, it follows that either the saturation function is logarithmic or there must be a gain control mechanism preceding the saturation mechanism controlling the present data. Table 1 illustrates that the saturation functions are far better fit by the hyperbolic function than by the logarithmic function. The present data, therefore, are incompatible with the absence of gain control preceding the measured nonlinearity.

The most appealing hypothesis is that the nonlinearity we measure is at the photoreceptor output. An early site of the nonlinearity is implied by the recent results of MacLeod *et al.* (1992), showing that distortion products for laser interference fringes of moderate luminance are visible up to spatial frequencies beyond the acuity limit, implying that the aperture for the distortion mechanism lies within the single cone pathway. They also show that the sinusoidal distortion product itself does not appear distorted, implying that there is no further distortion beyond the one with the single cone aperture.

A specific hypothesis for the saturation mechanism is a saturation in the number of channels available to carry the current in the outer segment membrane of the photoreceptors. This nonlinearity often has been described by the hyperbolic function of equation (3) (Naka & Rushton, 1966) but has more recently been fit for monkey and human cones by the exponential saturation function of equation (5) (Baylor *et al.*, 1987). In either case, the data are close to the current results. However, if the measured saturation is located at the photoreceptor output, the problem is to explain how the lateral effects in the gain clamping paradigm operate to clamp the gain mechanisms within the receptor outer segments.

Again, there are two possibilities. One is that the lateral gain control operates *via* the direct electrical coupling between cones. This is a feasible mechanism, but it requires that the gain clamping influences travel from the cone pedicle, where the direct coupling occurs, along the cone axon and through the inner segment to all regions of

the outer segment. This is a long reach for a clamping process to maintain, especially considering that the cone axon may be as long as 100 m in the foveola, where these experiments were conducted.

The other possibility is that the gain control mechanism operates by means of chemical transmission through the inter-receptor space between the invaginations of the outer segments. In this way, it could have direct access to the outer segment transduction mechanisms but also exhibit the requisite lateral pooling effects between cones.

An alternative to a photoreceptor nonlinearity is to suppose that a transduction gain control operates to ensure that the cones are never pushed to their limits under our stimulating conditions, and that the measured saturation function is that of the subsequent retinal neurons. In this case, the gain control could be at a post-receptoral stage for which the lateral interaction pathways are well known. This approach provides a mechanism for the gain clamping, and is also compatible with the evidence for a lack of gain control in human cone responses. The problem is now to explain the wide range of intensity gain control that seems to be available in the human light response. Presumably, the gain would be set in a post-receptoral neuron and be subject to lateral influences, but changes in time course with intensity have never been demonstrated in post-receptoral neurons.

REFERENCES

Baylor, D. A. & Hodgkin, A. L. (1974). Changes in time scale and sensitivity in turtle photoreceptors. *Journal of Physiology, 242*, 729–758.

Baylor, D. A., Nunn, B. & Schnapf, J. L. (1987) Spectral sensitivity of cones of the monkey *Macaca fascicularis*. *Journal of Physiology, 390*, 145–160.

Bowen, R. W., Pokorny, J. & Smith, V. C. (1989). Sawtooth contrast sensitivity: Decrements have the edge. *Vision Research, 29*, 1501–1509.

Boynton, R. M. & Whitten, D. N. (1970). Visual adaptation in monkey cones: Recordings of late receptor potentials. *Science, 170*, 1095–1111.

Buss, C. M., Hayhoe, M. M. & Stromeyer, C.F. III (1982). Lateral interactions in the control of visual sensitivity. *Vision Research, 22*, 693–709.

Daly, S. J. & Normann, R. A. (1985). Temporal information processing in cones: Effects of light adaptation on temporal summation and modulation. *Vision Research, 25*, 1197–1208.

de Lange, H. (1958). Research into the dynamic nature of the human fovea–cortex systems with intermittent and modulated light. I. Attenuation characteristics with white and colored light. *Journal of the Optical Society of America, 48*, 777–785.

Fuortes, M. G. H. & Hodgkin, A. L. (1964). Changes in time scale and sensitivity in the ommatidia of limulus. *Journal of Physiology, 172*, 239–263.

Graham, C. H. & Kemp, E. H. (1938). Brightness discrimination as a function of the duration of the increment in intensity. *Journal of General Physiology, 21*, 635–650.

Graham, N. & Hood, D. C. (1992). Modeling the dynamics of light adaptation: The merging of two traditions. *Vision Research, 32*, 1373–1393.

Hood, D. C. & Finkelstein, M. S. (1986). Sensitivity to light. In Boff, K., Kaufman, L. & Thomas, J. (Eds), *Handbook of perception and human performance* (Vol. 1, pp. 5-1–5-66). New York: Wiley.

Hood, D. C., Ilves, T., Maurer, E., Wandell, B. & Buckingham, E. (1978). Human cone saturation as a function of ambient intensity: A test of models of shifts of dynamic range. *Vision Research, 18*, 379–390.

Ikeda, M. (1965). Temporal summation of positive and negative flashes in the visual system. *Journal of the Optical Society of America, 55*, 1527–1534.

Kelly, D. H. (1961). Visual responses to time-dependent stimuli. I. Amplitude sensitivity measurements. *Journal of the Optical Society of America, 51*, 422–429.

Klein, S. A. & Tyler, C. W. (1993). The psychophysics of detection: A review of Graham's "Visual Pattern Analyzers". *Journal of Mathematical Psychology, 37*, 119–135.

MacLeod, D. I. A., Williams, D. R. & Makous, W. (1992). A visual nonlinearity fed by cones. *Vision Research, 32*, 347–364.

Naka, K. I. & Rushton, W. A. H. (1966). S-potentials from colour units in the retina of fish (*Cyprinidae*). *Journal of Physiology, 185*, 536–555.

Normann, R. A. & Perlmann, I. (1979). The effects of background illumination on the photoresponses of red and green cones. *Journal of Physiology, 286*, 491–507.

Rashbass, C. (1970). The visibility of transient changes of luminance. *Journal of Physiology, 210*, 165–186.

Schnapf, J. L., Kraft, T. W. & Baylor, D. A. (1987). Spectral sensitivity of human cone photoreceptors. *Nature, 325*, 439–441.

Schnapf, J. L., Nunn, B. J., Meister, M. & Baylor, D. A. (1990). Visual transduction in cones of the monkey *Macaca fascicularis*. *Journal of Physiology, 427*, 681–713.

Shapley, R. & Enroth-Cugell, C. (1986). Visual adaptation and retinal gain controls. In Osborne, N. N. & Chader, G. J. (Eds), *Progress in retinal research* (Vol. 3, pp. 263–346). Oxford: Pergamon Press.

Stiles, W. S. (1961). Adaptation, chromatic adaptation, colour transformation. *Anales de la Real Sociedad de Fisica y Quimica, 57A*, 149–175.

Tyler, C. W. (1991). Tacit assumptions in visual psychophysics. In Gorea, A. (Ed.), *Representations of vision: Tacit asssumptions in visual perception* (pp. 251–278). Cambridge, U.K.: Cambridge University Press.

Tyler, C. W., Chan, H. & Liu, L. (1992). Different spatial tunings for ON and OFF pathway stimulation. *Ophthalmic and Physiological Optics, 12*, 233–240.

Tyler, C. W. & Hamer, R. D. (1987). Analysis of visual modulation sensitivity. III. Meridional variations in peripheral flicker sensitivity. *Journal of the Optical Society of America A4*, 1612–1619.

Weibull, W. (1951). A statistical distribution function of wide applicability. *Journal of Applied Mechanics, 18*, 292–297.

Williams, T. S. & Gale, J. G. (1977). A critique of an increment threshold function. *Vision Research, 17*, 881–882.

Acknowledgements—Supported by NSF Grant NSF BNS 9011837 to CWT and a Rachel C. Atkinson Postdoctoral Fellowship to LL.

APPENDIX

In the Weibull function

$$W(c) = 1 - (1 - \gamma)e^{-\left(\frac{c}{\alpha}\right)\beta} \quad \text{(A1)}$$

when the signal contrast c equals the threshold contrast α, the proportion-of-seeing is

$$P(\alpha) = W(\alpha) = 1 - (1 - \gamma)/e \quad \text{(A2)}$$

If we fit the central portion of the proportion-of-seeing curve with a straight line with slope β_0 and γ-intercept A, then it is easy to show that the estimated threshold α_0 is

$$\begin{aligned} \alpha_0 &= [P(\alpha) - A]/\beta_0 \\ &= [1 - (1 - \gamma)/e - A]/\beta_0 \end{aligned} \quad \text{(A3)}$$